

Flexible Diffusion Modeling of Long Videos

Year: 2022 | Citations: 341 | Authors: William Harvey, Saeid Naderiparizi, Vaden Masrani, Christian Weilbach, Frank Wood

Abstract

We present a framework for video modeling based on denoising diffusion probabilistic models that produces long-duration video completions in a variety of realistic environments. We introduce a generative model that can at test-time sample any arbitrary subset of video frames conditioned on any other subset and present an architecture adapted for this purpose. Doing so allows us to efficiently compare and optimize a variety of schedules for the order in which frames in a long video are sampled and use selective sparse and long-range conditioning on previously sampled frames. We demonstrate improved video modeling over prior work on a number of datasets and sample temporally coherent videos over 25 minutes in length. We additionally release a new video modeling dataset and semantically meaningful metrics based on videos generated in the CARLA autonomous driving simulator.